基于意见领袖的微博生命周期预测模型研究 *

黄贤英，杨林枫?，刘小洋，高民东

(重庆理工大学 计算机科学与工程学院，重庆 400054)

摘　要：为了有效研究社交网络中意见领袖在新浪微博传播所起的作用及微博的生命周期和传播模式，提出了一种OLB 微博传播预测模型。首先通过爬取微博数据，进行数据分析；其次，拟合出与影响力有关的四个因素的数学表达式，并通过层次分析法给出权重计算方法；最后利用计算的影响力以及转发数与相关因素的关系构建出 OLB 模型，从而对意见领袖传播作用及微博生命周期进行实验预测分析。仿真结果表明，在微博信息传播中意见领袖影响力与其微博的传播作用成正比例关系，通过误差分析得到四组数据的平均误差值分别为 1.0%、5.0%、2.4%及 5.1%，提出的 OLB 模型对于预测微博传播模式合理、有效。

关键词：传播模型；意见领袖；影响力；微博生命周期

中图分类号：TP393　　doi: 10.3969/j.issn.1001-3695.2017.09.0945

Research on life cycle prediction model of micro-blog based on opinion leaders

Huang Xianying, Yang Linfeng?, Liu Xiaoyang, Gao Mindong

(*College of Computer Science & Engineering*, *Chongqing University of Technology*, *Chongqing* 400054, *China*)

Abstract: In order to study the role of opinion leaders in the dissemination of Sina micro-blog and the propagation mode of micro-blog's life cycle in social networks effectively, the paper proposed a propagation prediction OLB model. Firstly, we crawled the micro-blog data, then analyzed the data; Secondly, we fitted a mathematical expression of the four factors related to influence, and gave the weights calculation method through the analytic hierarchy process; Finally, we constructed the OLB model by using the influence of computation and the relationship between forwarding number and related factors. We analyzed the communication role of opinion leaders and micro-blog life cycle through the experiment. The simulation results showed that the influence of opinion leaders was positively proportional to the spread of micro-blog in micro-blog information dissemination. The average error of the four sets of data were 1.0%, 5.0%, 2.4% and 5.1% respectively by error analysis. The OLB model is reasonable and effective for predicting micro-blog propagation patterns.

Key Words: Communication model; Opinion leader; Influence; Micro-blog life cycle

0　引言

微博的兴起改变了人们的生活，作为一种新兴社交网络媒体，其特点是文本短小，内容精炼，信息传播速度快。近年来，人们获取新闻的第一途径已经逐渐转变为微博。微博在社交网络中扮演十分重要的角色[1,2]。而意见领袖作为信息传播当中的媒介，使信息传播形成二级传播，通过意见领袖造成信息传播的社会舆论也存在重大影响，引起更多的研究者关注，故对微博意见领袖影响力的以及微博生命周期的建模与分析是在线社交网络研究当中最前沿的问题之一。

近些年国内外诸多研究学者都对在线社交网络、复杂网络、微博信息传播等进行深入研究。针对意见领袖，对于识别意见领袖尤为重要。吴渝等人[1]通过对提出一种新的基于用户影响力的 PageRank 意见领袖的识别方法，发现意见领袖是随时间的变化而变化，只有很少的用户在不同的时间段一直是意见领袖；张树森、梁循等人[2]根据角色是否提前定义，将社会网络角色分为非明确和明确角色，发现社会网络角色的识别需要用“组合拳”方式来解决，识别出社会角色；Li 等人[3]提出一个框架，通过检索内容分析微博因素来识别标志的意见领袖和最大化的信息传播；吴岘辉等人[4]提出一种基于用户行为网络的意见领袖挖掘算法．并采用微博内容和社交属性，挖掘微博中的意见领袖。在识别过程中，运用算法作为识别方法进行识别。对意

基金项目：国家教育部人文社科青年基金资助项目（16YJC860010）；国家社科基金资助项目（17XXW004）；重庆市教委人文社会科学研究项目（17SKG144）；重庆市社会科学规划博士项目（2015BS059）

作者简介：黄贤英（1967-），女，重庆人，教授，硕导，主要研究方向为计算机应用、传播模型等；杨林枫（1993-），男（通信作者），陕西人，硕士研究生，主要研究方向为在线社交网络信息传播模型（lfyang@2016.cqut.edu.cn）；刘小洋（1980-），男，安徽人，副教授，硕导，博士（后），主要研究方向为传播模型、无线通信与计算机应用等；高民东（1994-），男，贵州人，硕士研究生，主要研究方向为计算机应用．

见领袖的识别工作逐渐成熟，而对意见领袖的传播方式却研究较少。

为了对意见领袖在原创或转发后的微博生命周期进行分析，首先要构建意见领袖传播力以及影响力因素指标。在在这之前需要获取微博数据，Deng 等人[5]构建了一个基于新浪微博 API 的微博爬行分析工具。这种基于微博客的新的 MapReduce 得到更好的精度和更少的运行时间检测微博客核心意见领袖；李玉贞等人[6]提出以活跃度、传播力和覆盖度三个指标来评估微博意见领袖的影响力；赵惠东等人[7]利用直接粉丝与间接粉丝构造影响力计算方法，并建立模型用以预测微博转发量；张仰森等人[8]采用层次分析提出了用户初始影响力的计算模型。并借鉴 PageRank 算法思想，提出了用户初始影响力扩散的计算方式；冯时等人[9]提出了一种基于 LDA 语义信息和 HowNet 知识库的短文本子话题分类算法。对分类后的微博从显式、隐式及用户等方面综合衡量微博的影响力；丁兆云等人[10]提出了基于多关系网络的随机游走模型 MultiRank，将影响力个体根据其影响力属性划分，结果表明多话题层次影响力个体数目相对所有影响力个体仅占少部分。而对于微博内容的分析也十分有必要，董雨辰，刘琰等人[11]利用社团模块度、平均最短路径和网络直径这三个属性度量该网络的紧密程度，基于支持向量机对所抽取的微博进行分类,进而识别出炒作微博。将机器学习算法与微博分类结合起来可以对微博进行分类[9][11]。对于以上研究均可作为对一条突发事件微博的生命周期进行分析的预处理工作。

而对于微博传播模式的研究及微博传播方式预测同样非常重要。Wei 等人[12]提出三种信息发布模式并基于构造了三种不同的扩散模型；Tang 等人[13]提出一种基于兴趣的动态交友网络，并构建了一种扩散模型；Zhang 等人[14]提出了一种新的方法来检测突发事件，并预测他们未来的发展；Jalali 等人[15]基于经验数据，提出了一种 DY 动态模型并利用最大似然估计给出模型的测试和校准方法；Long 等人[16]提出了一种新的基于定向网络的重叠社区标志算法结果表明：本研究中的社区识别算法实现了有向图到加权无向图的转换，扩展了整个社区的重叠网络社区。对于传播模式的研究中，没有用数学建模的思想考虑微博生命周期，同时也没有考虑在传播过程中意见领袖所起的作用。

从国内外相关研究学者对在线社交网络传播中意见领袖所起的作用研究表明传播过程中意见领袖具有重要意义及作用。微博的发展速度迅猛，近年来微博及各类社交网络软件信息传播量大，传播力度广，且人们通过微博获取日常新闻是一种普遍现象，故对社交网络研究。本文将基于新浪微博对其中各因素进行指标分析，计算相关因素的权重，分析出意见领袖影响力。并构建数学模型分析微博生命周期。

1 数据采集

新浪微博作为国内最大的网络媒体，也是全球使用最多的微型博客提供商之一。产生的数据量巨大，用新浪微博对在线社交网络行为进行分析具有重大意义。目前，新浪微博月活跃用户平均已达 3 亿，而日活跃用户平均也有 1 亿以上。

微博原创发送及微博转发都是信息传播行为，本文将会研究微博生命周期行为特征，主要是一条微博发送或转发之后，经历潜伏期、激增期、二次增长期最后到达衰亡期的规律特征，以意见领袖作为研究对象，分析其在发送或转发微博之后的微博传播模式。

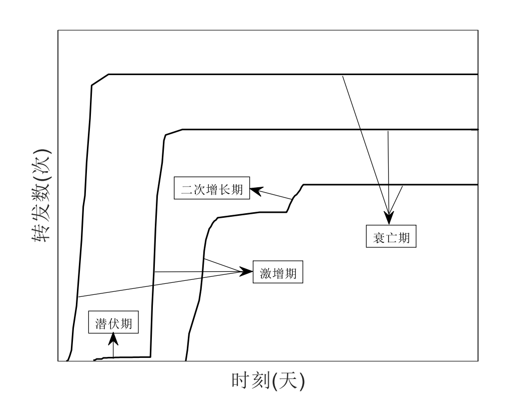


图 1　微博生命周期示意图

从图 1 可清晰看出一条微博的生命周期所经历的各个阶段，三条曲线分别描述了三类典型微博的传播模式。

本文所用到的数据将采用爬虫方式爬取数据，由于微博对用户信息有所保护，在爬取过程中，如果在某个时间段一个用户对微博页面访问量太大，将会封号，故爬取过程利用三个账号，分别爬取相关信息。首先对微博名人堂中的 200 个用户进行爬取，分别爬取每个用户的总微博量、原创微博量、关注列表数、影响力大小以及粉丝数量；并爬取每个用户的 50 篇微博用来计算其平均评论量以及平均转发量，而活跃度通过微博用户总微博量、平均每天发表微博数以及活跃天数来拟合。对爬取的 200 个用户进行一次筛选，去掉评论数及转发数极低的以及重复的人，得到 172 名用户。其次对微博的热搜数据进行爬取，爬取时间从 2017 年 3 月 20 日到 2017 年 6 月 30 日每天任选两个时间点。每个时间点只取 8 个热门话题，两个月共除去重复话题共取到 508 个热门话题。其中每个话题涉及的微博数目参差不齐，其中也有普通大众为了蹭热度而发布的与话题毫无关系的微博，故每个话题只取评论转发较高的 1~3 篇分析，最后共取到 822 篇微博。其中涉及的意见领袖共 1183 人。下面将对采集到的因素数据分析研究，并对热搜微博数据进行分析建模研究。

2 意见领袖影响力分析研究

进行相关数据采集后，将微博各因素数据归类，并对意见领袖影响力进行研究。首先对微博的传播模式进行分析，其次将会对于爬取到的与意见领袖影响力有关的各个指标，对其分析。最后构建意见领袖影响力权重计算方式。指标分类如表 1 所示。

表 1　意见领袖影响力指标分类

意见领袖传播力指标	意见领袖活跃度指标
粉丝数	发表微博总数
微博被评论数	平均每天发表微博数
微博被转发数	活跃天数

若仅考虑微博的意见领袖传播力，并不能准确计算和分析其影响力；则引入发表微博总数、平均每天发表微博数、活跃天数这三个因素来构建意见领袖活跃度，如表 1 所示；意见领袖活跃度与此三个因素均成正比关系。

2.1　意见领袖相关因素分析

对爬取到的数据。将影响力通过式(1)进行归一化处理。

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

用意见领袖转发微博的总数量;意见领袖关注列表的数量;意见领袖粉丝数量；意见领袖微博平均评论数量；意见领袖微博转发数量；意见领袖活跃度分别与意见领袖影响力进行绘图分析以发现因素的相关关系。

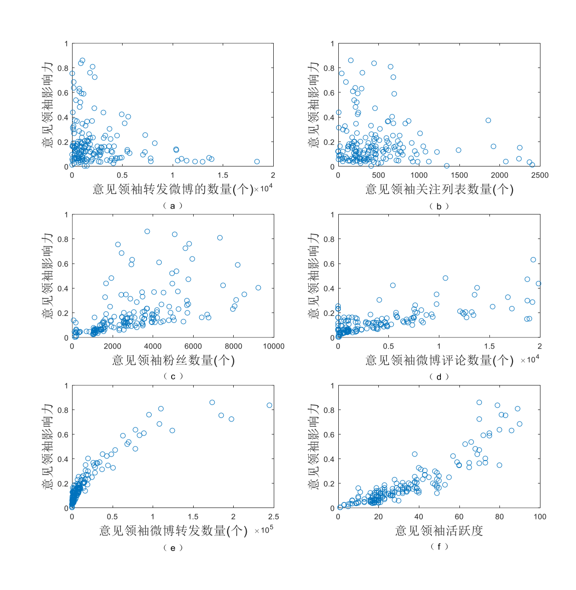


图 2　不同因素与意见领袖影响力关系

根据图 2 可以看出意见领袖所转发微博的数量以及其关注列表的数量对意见领袖的影响力不存在函数关系，对其他四个因素用式(1)将其数值进行归一化至区间[0,1]，然后分别用 V_{fans}、$V_{comment}$、$V_{forward}$ 以及 $V_{activity}$ 来表示。并对图 2 中其他四个因素对影响力的关系进行函数拟合。

由图 3 可以看出四个因素对影响力均存在函数关系，通过拟合出的四条线性直线可得式(2)~(5)。

$$V_{fans} = 0.469x_{fans} + 0.042 \tag{2}$$

$$V_{comment} = 1.081x_{comment} + 0.118 \tag{3}$$

$$V_{forward} = 0.684x_{forward} + 0.084 \tag{4}$$

$$V_{activity} = 0.691x_{activity} - 0.0592 \tag{5}$$

通过式(2)~(5)可得意见领袖粉丝数、微博评论数、转发数以及意见领袖的活跃度四个因素确实与意见领袖影响力存在函数关系，接下来分析拟合曲线的误差及各指标值。

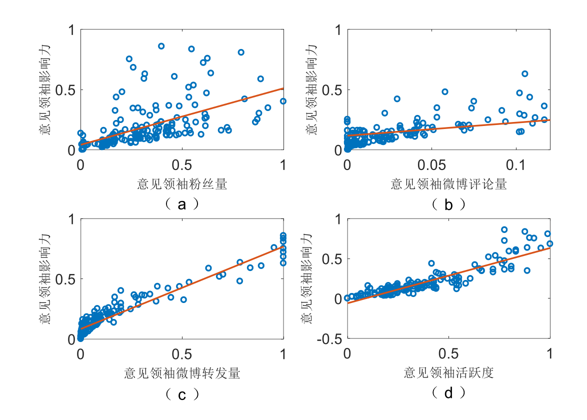


图 3　各因素与意见领袖影响力拟合关系图

根据表 2 中的各指标首先可以分析出四个因素与影响力都存在线性关系，其次发现与影响力关系最大的指标是微博转发量。关系最小的是意见领袖的粉丝数量。

表 2　拟合曲线指标分析表

指标	SSE	R-square	Adjusted R-square	RMSE
粉丝量-影响力	2.811	0.4128	0.4087	0.1297
评论量-影响力	1.894	0.6584	0.6564	0.1056
转发量-影响力	0.4294	0.9226	0.9221	0.05026
活跃度-影响力	1.055	0.8097	0.8085	0.07879

2.2　意见领袖影响力构建与分析

微博在通过意见领袖传播的过程中，对其传播作用起到最大影响的是意见领袖的影响力，因影响力这个指标是由多个因素共同构成，所以影响力对于信息的传播起到关键性作用，即意见领袖影响力越大，其带来的网络传播力度越广，所造成的社会影响越大。故将利用意见领袖粉丝数、微博评论数、转发数及活跃度结合式(6)来计算意见领袖影响力。意见领袖影响力用 f 表示。

$$f = w_1V_{fans} + w_2V_{comment} + w_3V_{forward} + w_4V_{activity} \tag{6}$$

其中：活跃度是由微博总数、平均每天发表微博数以及活跃天数三因素分别归一化并相加取平均得到。

接着将确定各因素在影响力计算中的权值，对属性矩阵数据归一化处理可以在[0,1]范围内比较数据，但各属性值在评价中的重要性不能确定，因此还需要确定属性的权重值。采用 Saaty 的 10 级重要性等级表和两两比较的判断矩阵专家法计算各个属性指标的相对重要性。比如，评价人认为意见领袖粉丝数量 x_1 与微博转发数 x_2 和意见领袖活跃度 x_3 相比相当重要，则可取 $b_{12} = b_{13} = 5$；相比较明显可以取 7；介于相当重要和明显重要之间可以取 6；介于略微重要与相当重要之间可以取 4。依次对各个指标进行比较后，运用层次分析法，则得到属性相对重要性判断矩阵。最后通过一致性检验来确定各因素权重数据，如表 2 所示。

表 3　意见领袖影响力因素权重

因素	粉丝数量	平均评论数	平均转发数	活跃度
权重	0.22	0.12	0.36	0.30

通过表 3 中的权重来计算意见领袖的影响力，从而更好的分析微博生命周期。(无特殊情况外，转发量高于 10 万的微博都认为有水军参加，忽略这些转发量，本文将微博实际转发量超过 10 万的定义为 10 万)。

3　提出的意见领袖 OLB 传播模型

3.1　微博生命周期分析

从数据采集的 822 篇微博中随机挑选 10%即 82 篇微博进行分析，可以看出一条热门微博的生命周期，主要是两种传播模式，如图 4 所示。图 4 的生命周期与图 1 的生命周期有所区别，主要是新浪微博因其实时性较高，在其发布的第一时间内就会得到普通大众及意见领袖的传播，故在其生命周期中很少会存在潜伏期阶段。

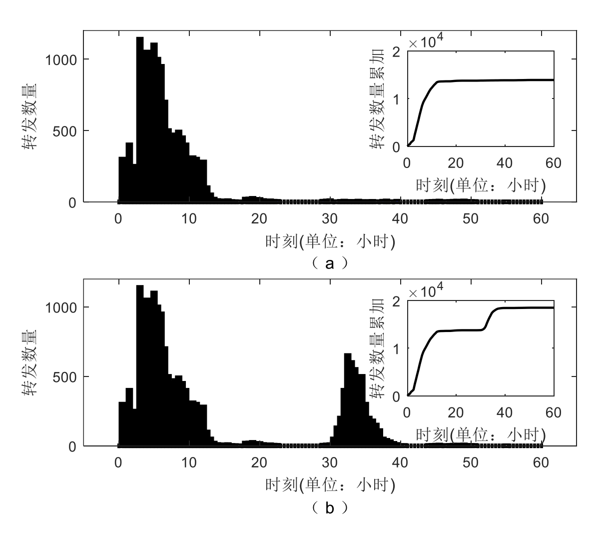


图 4　一条典型微博的传播模式图

一条微博的生命周期主要由潜伏期、激增期、二次增长期、衰亡期构成。图 4 中的黑色阴影部分是以每小时微博的转发量画图所得，图 4 右上角的曲线部分是将每个小时的转发量累加起来所画出的图，其可以描述一条典型微博的两种不同的传播模式。图 4(a)中是微博再转发之后由激增到达衰亡的过程，通过右上角小图可以明显看到激增期以及衰亡期两个阶段；而图 4(b)相对于图 4(a)多了一个二次增长的阶段，二次增长主要是由于在微博转发数量已经达到平稳阶段后又被一影响力大的意见领袖进行转发导致，微博转发数进行二次增长。

3.2　OLB 模型建立

一般情况下，意见领袖的影响力在信息传播过程中在逐步衰减，本文中，意见领袖影响力由 2.1 节和 2.2 节所提出来的四个因素构建。并且假定在信息传播过程中，粉丝数 n_{fans}、微博评论数 $n_{comment}$、微博转发数 $n_{forward}$ 以及活跃度 $n_{activity}$ 不会发生变化。

意见领袖影响力 f 的大小可以影响信息传播的力度，而一条微博的转发数主要跟信息的传播力度有关，信息的传播力度越大，就能使更多人看到，所造成的微博转发数也就越大。意见领袖影响力根据式(2)~式(5)结合权重计算式(7)进行计算，下一步将判断计算出的意见领袖影响力 f 跟相应的转发数 $N_{forward}$ 存在什么样的函数关系，对两个因素进行拟合分析得到图 5。

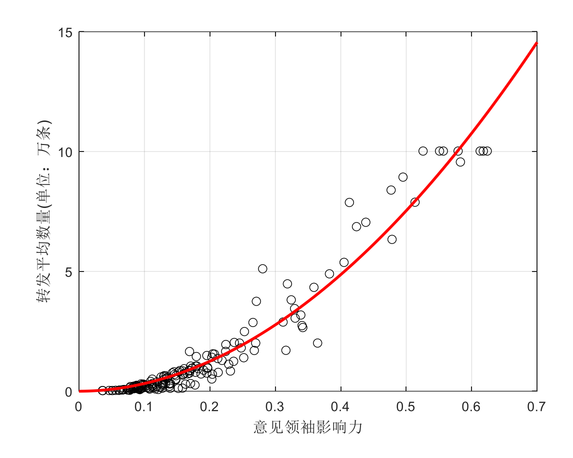


图 5　影响力与转发数关系曲线

由图 5 可以看出意见领袖平均微博转发数与其影响力存在函数关系。且保持正比例递增。拟合得到的 R-square 值为 0.9473，说明拟合结果有效。

$$N_{forward} \propto f \tag{7}$$

$$N_{forward} = 29.3 f^2 \tag{8}$$

一条微博的激增期，转发数会随时间 t 的增加而增大，但转发数最终会达到一个稳定不变的状态，把转发数增长的这段时间称为影响力持续时间 t_c，当 $t < t_c$ 时，微博转发数在激增，可用式(9)来拟合出转发数在激增过程中的大体变化趋势，t_k 作为时间控制参数。

$$y = 1 - e^{(t_k - t)} \tag{9}$$

对微博转发数的总值在模型中引入放大级数 β，放大级数是对数学模型计算出的转发数进行实际放大，发现放大级数与意见领袖影响力存在关系，其增大方式是影响力越大，放大级数 β 也就越大，函数关系如式(10)所示。

$$\beta = \begin{cases} 10f & 0 < f < 0.08 \\ 1.06f + 0.716 & 0.08 \le f < 0.25 \\ 34.375f - 8.2125 & 0.25 \le f < 0.35 \\ 11.165f - 0.878 & 0.35 \le f < 1 \end{cases} \tag{10}$$

当 $t < t_c$ 时，微博生命状态处于激增阶段，此时微博转发数随时间快速增长，如式(11)所示。

$$N_{forward} = 10f^2 + \beta(1 - e^{2(t_k - t)}) \tag{11}$$

当 $t = t_c$ 时，微博生命状态处于激增结束阶段，开始缓慢增加至稳定态。此时可以根据式(11)计算出最终转发数。

对计算出的转发数，需对模型转发数 $N_{forward}$ 与 N_{real} 进行误差分析，利用式(12)~(13)进行误差计算并进行分析。

$$\delta = \frac{\left|N_{forward} - N_{real}\right|}{N_{real}} \tag{12}$$

$$\bar{\delta} = \frac{\delta_1 + \delta_2 + \cdots + \delta_n}{n} \tag{13}$$

根据计算出的误差平均值 $\bar{\delta}$ 的大小可以分析 OLB 模型的有效性以及是否具有实际意义。

4 仿真结果与分析

本文仿真数据利用热门微博进行仿真，构建的微博意见领袖预测模型是针对意见领袖发送的原创微博有效。而由于一般微博评论量及转发量均具有随机性，则提出的预测模型适用于意见领袖的热门微博。

4.1 微博热搜微博热度分析

在之前爬取的数据中选取四个不同类别且转发量多的话题，以拟合微博事件热度图，可以分析一个话题从出现并被人们关注的热度变化趋势。仿真前将数值归一化。

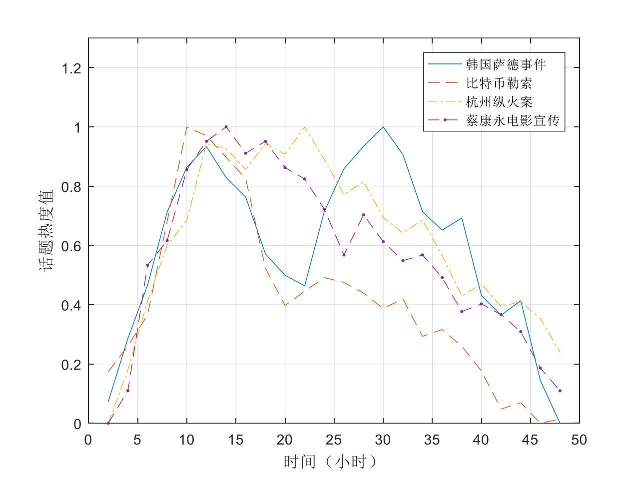


图 6 微博话题热度图

由图 6 可以看出一个话题在发布后，人们对其极为关注，热度也在攀升，但当时间逐渐增加，事件将会被人们淡忘，热度也会降低，各路媒体会在人们淡忘时爆出话题的附属事件即可引起更大关注量，使人们重新关注事件，事件的热度会再次增加，如此循环，最后热度会降到最低。

4.2 真实数据与提出的 OLB 模型对比分析

利用热度分析中的 4 个不同话题的微博进行分析，首先对意见领袖粉丝数及活跃度进行仿真分析，其次从转发的人当中寻找意见领袖，并计算意见领袖的影响力 f，在计算影响力时各因素均采集近一年内的数据，最后进行真实数据、传统模型方法以及 OLB 模型的对比仿真分析(其中用不同颜色代表不同的意见领袖)，其中采用文献[7]中的模型，此方法当前运用较多，采用时刻转发量进行预测，本文将其转发量累积起来进行对比，对比效果更直观。

表 4 蔡康永发布电影宣传微博参数

意见领袖	杨幂	王嘉尔	范冰冰	张大大
传播力 f	0.625	0.316	0.268	0.051
放大级数	6.10	2.65	1.00	0.51
转发间隔时间(天)	0.20	0.13	0.29	0.42
实际转发量(万)	12.35	3.66	1.70	0.54

如表 4 是针对蔡康永首部导演作品所发的宣传微博，各明星意见领袖都帮其转发宣传，起到信息传播的作用，其中筛选出杨幂、王嘉尔、范冰冰、张大大四个意见领袖进行模型与真实数据的仿真分析。

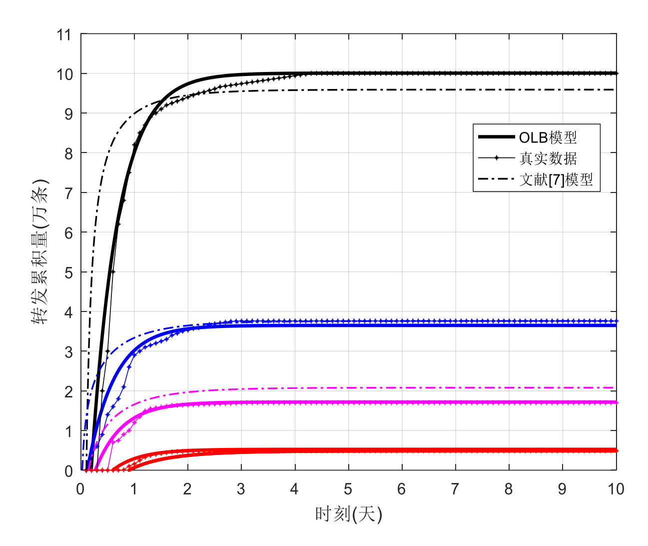


图 7 蔡康永发布电影宣传微博生命周期

据图 7 可以得到，意见领袖在转发别人微博的时间点没有规律。而特征都为先激增后衰亡，因为人们对信息的消化时间很快，在信息出现后，迅速关注令其热度的传播速度达到一个峰值又迅速下降至零，表现在微博累积转发量当中即为达到一个高点后不会再变化。并且从图中可以看出，杨幂对信息的传播作用力最大，其在娱乐圈有较大影响力。才导致在信息传播当中起较大作用，而单四个意见领袖的转发量就占总转发量的 57%，可见意见领袖在微博传播中的作用之大，真实数据与 OLB 模型误差平均值为 1.8%，与传统模型对比发现传统方法两个意见领袖的误差很大，其他两个接近。

表 5 乐天萨德”事件微博参数

意见领袖	最神奇的视频	我的损友是个极品	Happy 张江
传播力 f	0.281	0.080	0.035
放大级数	1.45	0.80	0.35
转发间隔时间(天)	0.13	0.23	1.05
实际转发量(万)	2.17	0.88	0.33

而表 5 是关于“乐天萨德”事件微博，涉及的意见领袖多为自媒体，微博自媒体是微博时代产生的一种通过微博传播信

息的网络媒体，其影响力在信息传播中也起很大的作用。

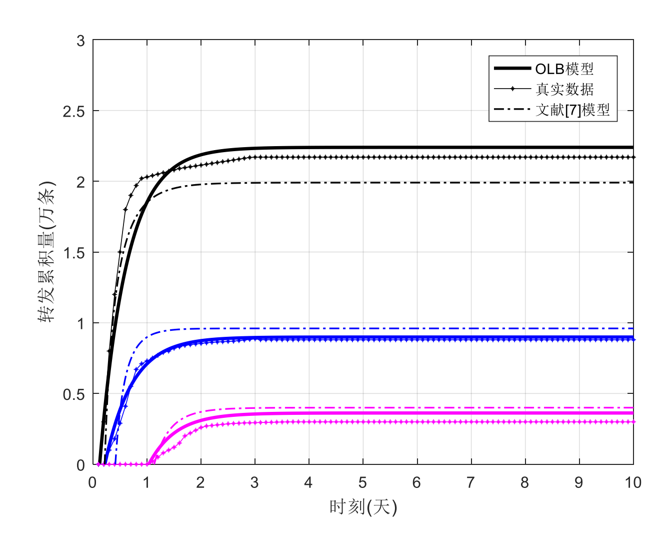


图 8　乐天萨德”事件微博生命周期

图 8 依旧在时间点上没有规律，而且关于这种政治方面的问题多为自媒体进行转发传播，一些知名人士不方便公开表态。关于传播模式还是先激增后衰亡，人们很快失去对信息的兴趣。而三个微博自媒体的转发量占总转发量的 59%，在传播过程中意见领袖起很大影响。而比较后的平均误差为 5.0%，与传统模型对比发现 OLB 模型对其中两个意见领袖表现更优秀。

表 6　比特币勒索病毒”解决方法微博参数

意见领袖	黑客凯文	夏目家的小诗哥	游研社
传播力 f	0.297	0.053	0.007
放大级数	1.65	0.53	0.07
转发间隔时间(天)	0.2	0.44	1.06
实际转发量(万)	2.52	0.59	0.07

表 6 是 3 月爆发的“比特币勒索病毒”事件，作为一种病毒软件，锁定各种文件，付费解锁。造成至少 150 个国家，30 万名用户中招、损失达 80 亿美元。与此话题相关的意见领袖多位 IT 行业意见领袖。用户“黑客凯文”发表中此病毒后如何解决的微博。

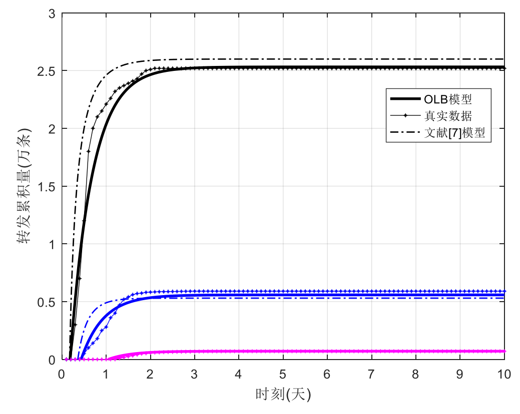


图 9　比特币勒索病毒”解决方法微博生命周期

如图 9 和前两组数据一样选三个有代表性的意见领袖进行仿真分析。可以看出其在 IT 行业中的传播模式依旧保持增长达到峰值即保持不变。此图中三个意见领袖造成的传播量占总传播量的 54%，意见领袖在微博信息传播当中依旧体现出很大的作用。而真实数据与 OLB 模型对比的误差平均值为 2.4%，OLB 模型中有两个意见领袖都优于传统模型。

表 7　杭州纵火案”微博参数

意见领袖	休闲璐	马思纯	夏目家的小诗哥
传播力 f	0.184	0.089	0.053
放大级数	0.91	0.79	0.53
转发间隔时间(天)	0.03	0.08	0.44
实际转发量(万)	1.3	0.83	0.61

表 7 数据为 6 月份震惊社会的“杭州纵火案”事件，此事件的影响之大导致涉及的意见领袖较多，选取的微博来自受害家庭的男主人所发送微博，因为事件过于恶劣，导致各方意见领袖：自媒体、明星以及新闻媒体都有参与。

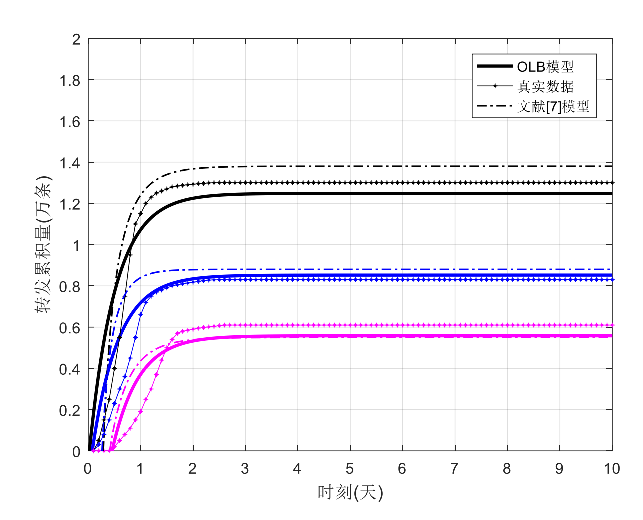


图 10　杭州纵火案”微博生命周期

图 10 中相对误差比前三组数据较明显，原因是微博当中的多级传播，在微博中可能一个影响力较小的意见领袖的微博被一个影响力较大的意见领袖所转发，将会导致前者的转发量与实际传播效果不成比例。上例中的真实数据与模型误差平均值为 5.1%，图中 OLB 模型针对三个意见领袖误差都小于传统模型。

利用图 7~10 可以得出 OLB 模型针对意见领袖进行微博预测优于传统模型。证明 OLB 模型合理有效，对于预测微博有意义。

4.3　误差分析

在对四组微博数据进行仿真分析后，得到 OLB 模型在预测意见领袖转发微博后的生命周期有着明显的有效性。进行误差计算之后得到四组微博的平均误差值分别为 1.0%、5.0%、2.4%以及 5.1%。对于每个人的误差与文献[7]中模型对比分析如图 11 所示。

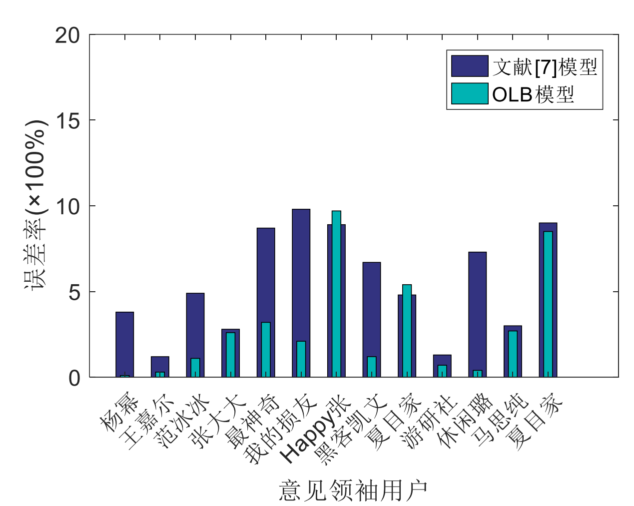


图 11　意见领袖模型误差分析

由图 11 可以看出，意见领袖影响力越大，提出的 OLB 模型的误差值越小；而意见领袖影响力变小的过程中，OLB 模型的误差值不稳定。相比较传统模型，本文模型对于大多数意见领袖误差都有明显提升。而总体 OLB 仿真得到的数据与真实数据相比误差较小，在微博实际当中模型可以较好的起到预测作用。

5　结束语

针对在线社交微博网络当中对于预测微博的生命周期没有确定的数学模型。本文对新浪微博中的微博特征进行预测研究，发现微博通过意见领袖的传播可以达到更大的效果，而转发量是导致一条微博扩散能力的关键因素，对与影响力有关的因素进行函数拟合得到意见领袖粉丝量、微博平均评论量、平均转发量以及其活跃度四个因素与意见领袖的影响力有线性关系；进而利用层次分析法对四个因素赋予相应的权重，得到意见领袖的相应影响力大小；在继续分析微博各因素间的相关关系的基础上构建了 OLB 数学模型。最后通过仿真及误差分析验证了提出的 OLB 模型在实际应用中的有效性。下一步本文将对微博内容进行分类、对意见领袖影响力的持续时间在模型中的界定问题进行继续研究。

参考文献：

[1] 吴渝，马璐璐，林茂. 基于用户影响力的意见领袖发现算法 [J]. 小型微型计算机系统, 2015, 36 (3): 561-565.

[2] 张树森，梁循，齐金山. 社会网络角色识别方法综述 [J]. 计算机学报, 2017, 40 (3): 649-673.

[3] Li F, Du T C. Maximizing micro-blog influence in online promotion [J]. Expert Systems with Applications, 2017, 70: 52-66.

[4] 吴岘辉，张晖，赵旭剑，等. 基于用户行为网络的微博意见领袖挖掘算法 [J]. 计算机应用研究, 2015, 32 (9): 2678-2683.

[5] Deng X, Li Y, Lin S. Parallel Micro Blog Crawler Construction for Effective Opinion Leader Approximation [J]. AASRI Procedia, 2013, 5 (3): 170-176.

[6] 李玉贞，胡勇，熊熙. 微博意见领袖的评估模型 [J]. 信息安全与通信保密, 2013 (2): 79-81.

[7] 赵惠东，刘刚，石川，等. 基于转发传播过程的微博转发量预测 [J]. 电子学报, 2016, 44 (12): 2989-2996.

[8] 张仰森，蒋玉茹，陈若愚. 微博用户关系网络中意见领袖的分析与挖掘 [J]. 北京信息科技大学学报, 2015 (4): 7-14.

[9] 冯时，景珊，杨卓. 基于 LDA 模型的中文微博话题意见领袖挖掘 [J]. 东北大学学报, 2013, 34 (4): 490-493.

[10] 丁兆云，周斌，贾焰. 微博中基于多关系网络的话题层次影响力分析 [J]. 计算机研究与发展, 2013, 50 (10): 2155-2175.

[11] 董雨辰，刘瑛，罗军勇. 基于支持向量机的炒作微博识别方法 [J]. 计算机工程, 2015, 41 (3): 7-14.

[12] Wei J, Bu B, Liang L. Estimating the diffusion models of crisis information in micro blog [J]. Journal of Informetrics, 2012, 6 (4): 600-610.

[13] Tang M, Mao X, Guessoum Z, et al. Rumor diffusion in an interests-based dynamic social network [J]. The Scientific World Journal, 2013, 3 (8): 2-9.

[14] Zhang X, Chen X, Chen Y, et al. Event detection and popularity prediction in microblogging [J]. Neurocomputing, 2015, 149 (8): 1469-1480.

[15] Jalali M S, Ashouri A, Herrera-Restrepo O, et al. Information diffusion through social networks: the case of an online petition [J]. Expert Systems with Applications, 2016, 44 (9): 187-197.

[16] Long H, Li B. Overlapping Community Identification Algorithm in Directed Network [J]. Procedia Computer Science, 2017, 107 (3): 527-532.